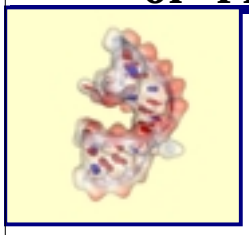
Structural Biology of Proteins & Nucleic Acids



Dr. Stephen R. Holbrook

Staff Scientist

Departments of Structural Biology
Computational & Theoretical Biology
Physical Biosciences Division
Lawrence Berkeley National Laboratory

Why Structural Biology?

- **Understand the Functions of Biological Macromolecules**
- **Determine Evolutionary Relationships**
- **Protein (and RNA) Engineering**
- **Ligand (Drug) Design**

Topics to be Covered

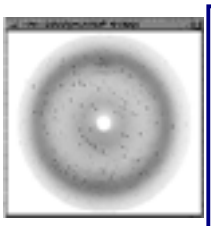
- Methods
- Databases
- Classification
- Visualization
- Modeling
- Molecular Interactions/ Docking
- Drug Design
- Summary

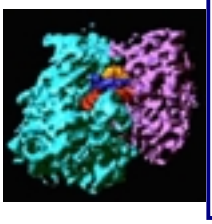
Methods for Determination of Three-Dimensional Structure

- Experimental
  - X-ray crystallography
  - Multi-dimensional NMR
- Computational
  - Molecular Modeling
  - Fold Prediction/Threading
  - *Ab initio*

X-Ray Crystallography of Biological Macromolecules

http://www-structure.llnl.gov/Xray/101index.html

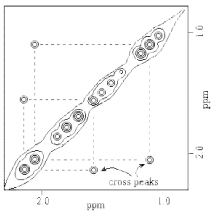




- **Advantages: Most powerful**
  - **Any size (ribosome, viruses)**
  - **High resolution/overdetermination**
  - **Waters, Metals, Ligands**
  - **Dynamics (Thermal parameters)**
- **Disadvantages**
  - **Need large quantities/highly purified**
  - **Must crystallize**
  - **Can be time consuming**

NMR Structure Determination of Biological Macromolecules

http://www.cryst.bbk.ac.uk/PPS2/projects/schirra/html/home.htm

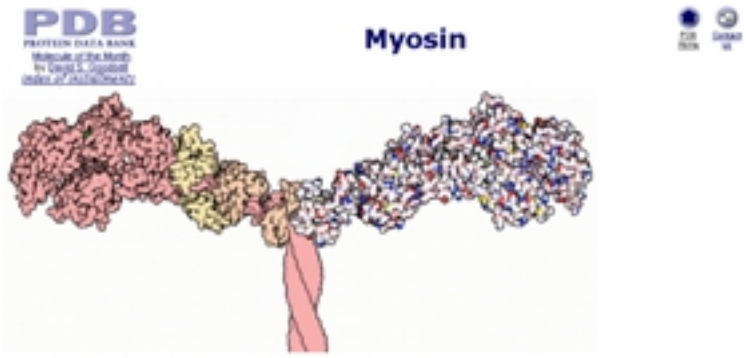


- **Advantages**
  - **Data collected in solution**
  - **Can easily change conditions**
  - **Do not need to crystallize**
  - **Dynamics**
- **Disadvantages**
  - **Need milligram quantities (isotopes)**
  - **Size limited to < 20-30 kDa**
  - **Local information, not global**
  - **Lower "resolution"**

Databases: RCSB Protein Data Bank

http://www.rcsb.org/pdb/index.html

Anatomy of a Molecular Muscle

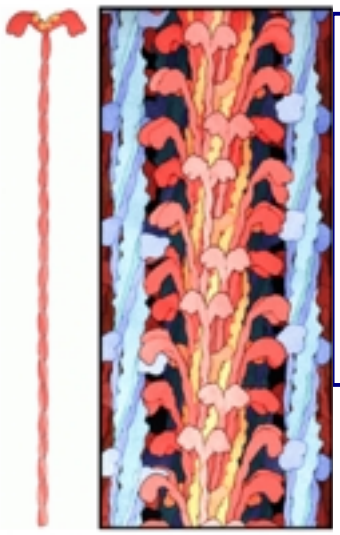


Molecular Motion

All of the different movements that you are making right now--your fingers on the computer keys, the scanning of your eyes across the screen, the isometric contraction of muscles in your back and abdomen that allow you to sit comfortably--are powered by myosin. Myosin is a molecule-sized muscle that uses chemical energy to perform a deliberate motion. Myosin captures a molecule of ATP, the molecule used to transfer energy in cells, and breaks it, using the energy to perform a "power stroke." For all of your voluntary motions, when you flex your biceps or blink your eyes, and for all of your involuntary motions, each time your heart beats, myosin is providing the power.

Myosin is composed of several protein chains: two large "heavy" chains and four small "light" chains. The structures available in the PDB, such as the one shown above, contain only part of the myosin molecule. In the illustration above, from PDB entry 1b7t, atoms in the heavy chain are colored red on the left-hand side, and atoms in the light chains are colored orange and yellow. The whole molecule is much larger, with a long tail that has been clipped off to allow the molecule to be studied. Fortunately, the crystal structures include most of the "motor" domain, the part of the molecule that performs the power stroke, so we can look at this process in detail.

Anatomy of a Molecular Muscle



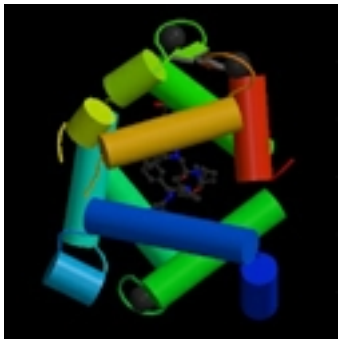
Power in Numbers

Each myosin performs only a tiny molecular motion. It takes about 2 trillion myosin molecules to provide the force to hold up a baseball. Our biceps have a million times this many, so only a fraction of the myosin molecules need to be exerting themselves at any given time. By working together, the tiny individual power stroke of each myosin is summed to provide macroscopic power in our familiar world. The painting shows how myosin is arranged inside muscle cells. About 300 myosin molecules bind together, with all of the long tails bound tightly together into a large "thick filament." A short segment of a thick filament is shown in red, next to a scale drawing of a single myosin molecule. The many myosin heads extending from the thick filament then reach over to actin filaments, shown in blue and green, and together climb their way up.

Databases: A Protein Data Bank File

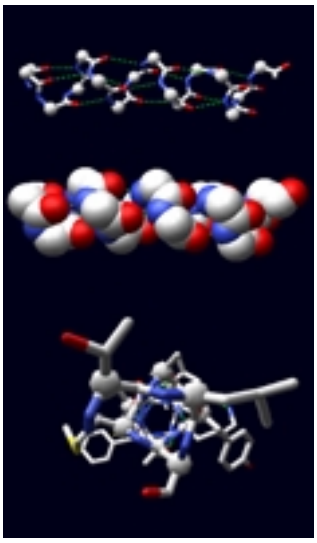
Search Keyword: Cardiac
31 PDB file hits

1DTL Deposited: 12-Jan-2000 Exp. Method: X-ray Diffraction Resolution: 2.15 Å
Title Crystal Structure Of Calcium-Saturated (3Ca2+) Cardiac Troponin C Complexed With The Calcium Sensitizer Bepridil At 2.15 A Resolution

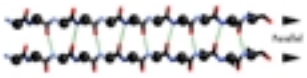


```
ATOM      1  N   TYR A   5      15.850  -8.428   2.760  1.00 41.77           N
ATOM      2  CA  TYR A   5      15.686  -9.919   2.617  1.00 42.13           C
ATOM      3  C   TYR A   5      15.079 -10.614   3.847  1.00 41.62           C
ATOM      4  O   TYR A   5      15.122 -10.097   4.971  1.00 41.33           O
ATOM      5  CB  TYR A   5      17.022 -10.583   2.258  1.00 54.00           C
ATOM      6  CG  TYR A   5      17.061 -11.068   0.834  1.00 56.45           C
ATOM      7  CD1 TYR A   5      16.046 -11.885   0.323  1.00 57.61           C
ATOM      8  CD2 TYR A   5      18.107 -10.720   0.000  1.00 57.88           C
ATOM      9  CE1 TYR A   5      16.081 -12.338  -0.982  1.00 59.55           C
ATOM     10  CE2 TYR A   5      18.154 -11.166  -1.305  1.00 60.05           C
ATOM     11  CZ  TYR A   5      17.140 -11.979  -1.797  1.00 60.72           C
ATOM     12  OH  TYR A   5      17.191 -12.449  -3.094  1.00 62.68           O
ATOM     13  N   LYS A   6      14.517 -11.792   3.617  1.00 39.40           N
ATOM     14  CA  LYS A   6      13.862 -12.533   4.676  1.00 38.68           C
ATOM     15  C   LYS A   6      14.698 -12.685   5.921  1.00 37.49           C
ATOM     16  O   LYS A   6      14.195 -12.489   7.038  1.00 36.25           O
ATOM     17  CB  LYS A   6      13.429 -13.905   4.166  1.00 82.36           C
ATOM     18  CG  LYS A   6      12.820 -14.819   5.228  1.00 82.36           C
ATOM     19  CD  LYS A   6      13.729 -16.007   5.495  1.00 82.36           C
ATOM     20  CE  LYS A   6      13.136 -16.935   6.543  1.00 82.36           C
ATOM     21  NZ  LYS A   6      14.017 -18.119   6.781  1.00 82.36           N
ATOM     22  N   ALA A   7      15.961 -13.041   5.729  1.00 41.54           N
ATOM     23  CA  ALA A   7      16.867 -13.217   6.851  1.00 40.69           C
ATOM     24  C   ALA A   7      17.090 -11.907   7.598  1.00 39.56           C
ATOM     25  O   ALA A   7      16.991 -11.877   8.838  1.00 37.44           O
```

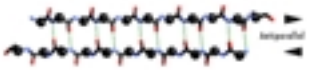
Elements of Protein Structure



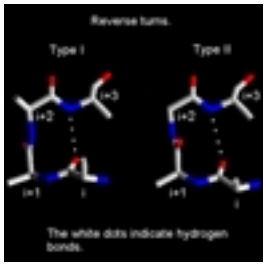
Alpha Helices



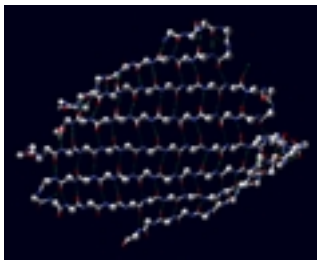
Beta Strands



Reverse Turns



Beta Sheet



Databases: NDB Nucleic Acid Database

http://ndbserver.rutgers.edu/

The Nucleic Acid Database Project (NDB) assembles and distributes structural information about nucleic acids

NDB News:
Files from data uniformity processing have been released. Details are available in the NDB Newsletter.

Search Search the Nucleic Acid Database, the NMR Nucleic Acid Database, and the DNA Binding Protein Database. Tutorials are available.

Atlas The Atlas of Nucleic Acid Containing Structures highlights the special aspects of each structure in the NDB using a combination of text and graphics. The Atlas has been used to create an Introduction to Nucleic Acids. A Musical Atlas is also available.

Archives The NDB Archives contain nucleic acid standards, nucleic acid summary information, software programs, NDB Newsletters, coordinate files, and structure factor data.

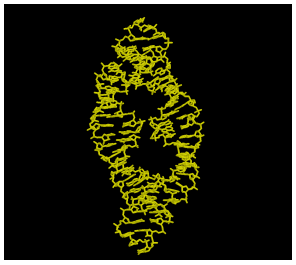
Deposition Data for the crystal structures of nucleic acids are deposited using the AutoDep Input Tool. Deposition information is also available.

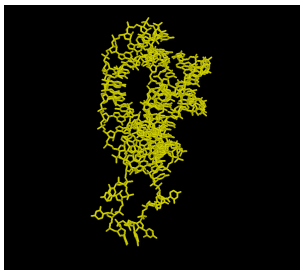
Resources The NDB maintains the mmCIF Web site, a Q&A mode, and links to other resources.

RNAs, Like Proteins, are Structurally Diverse

tRNA^{Phe}
6TNA

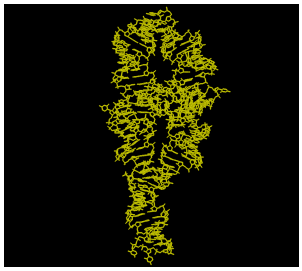
Hammerhead Ribozyme
1MME





HDV Ribozyme
1DRZ

P4-P6 Domain of Group I Intron
1GID



CLASSES: SCOP STRUCTURAL CLASSES

http://scop.berkeley.edu/

Structural Classification of Proteins

Structural Classification of Proteins

Welcome to SCOP: Structural Classification of Proteins. **1.53 release**
11410 PDB Entries (1 Jul 2000). 40 Literature References. 26219 Domains (excluding nucleic acids and theoretical models).
Folds, superfamilies, and families statistics here.

Authors: Alexey G. Murzin, Loredana Lo Conte, Bartlett G. Adley, Steven E. Brenner, Tim J. P. Hubbard, and Cyrus Chothia

MRC Laboratory of Molecular Biology and Centre for Protein Engineering
Hills Road, Cambridge, CB2 2QH, England: scop@mrc-lmb.cam.ac.uk

Reference: Murzin A. G., Brenner S. E., Hubbard T., Chothia C. (1995). SCOP: a structural classification of proteins database for the investigation of sequences and structures. *J. Mol. Biol.* 247, 536-540.

SCOP mirrors around the world may speed your access.

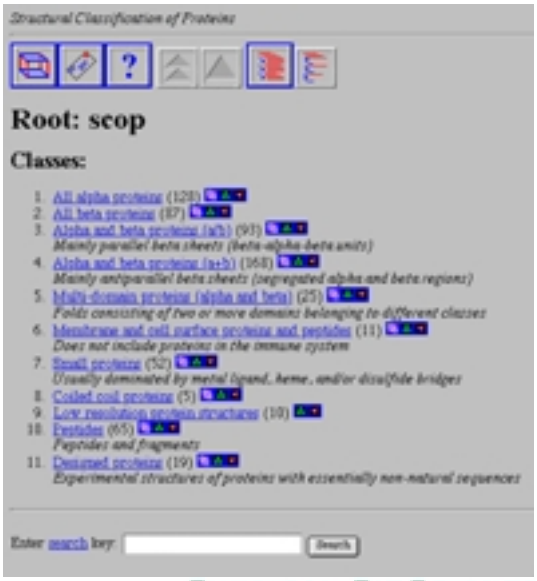
Access methods

- Enter SCOP at the top of the hierarchy
- Keyword search of SCOP entries
- PDB entry viewer
- SCOP sequence libraries and parseable files
- SCOP domain sequences (ASTRAL)

- Hidden Markov Model library for SCOP superfamilies (HYPERFAMILY)
- Intermediate Sequence Library (ISL)
- Online resources of potential interest to SCOP users

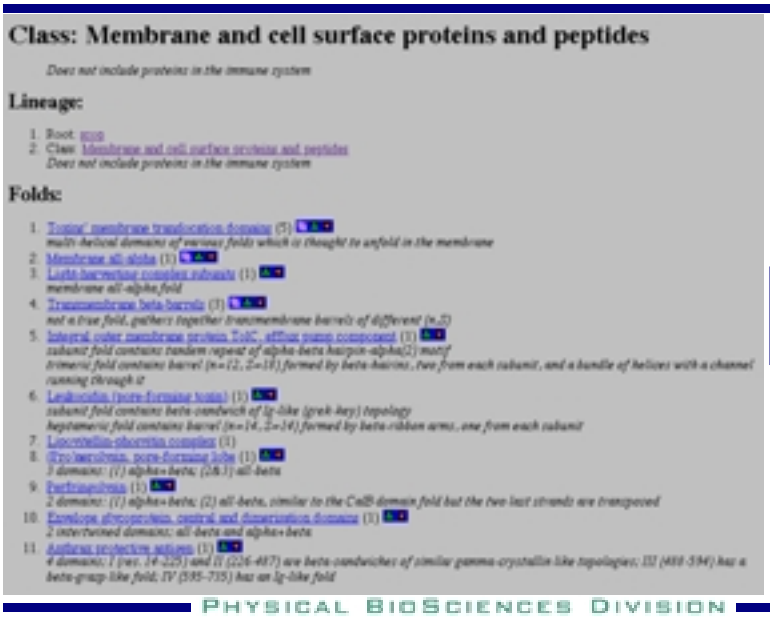
Warning: A new set of identifiers for all nodes in the SCOP hierarchy and corresponding parseable files will be introduced in the next release. If you are still referring to SCOP via classification page numbers (scop.html), please read more about reliably linking to SCOP first, and then don't hesitate to get in touch with us if you have any further questions.

SCOP STRUCTURAL HIERARCHY



- ✓ **Follow the structural hierarchy from the general to the specific**

SCOP STRUCTURAL HIERARCHY



- ✓ **This is the "Fold" Level**

SCOP Structural Hierarchy

Superfamilies are Specific

Fold: TIM beta/alpha-barrel

contains parallel beta-sheet barrel, closed; n=8, 2=8; strand order 12345678
the first six superfamilies have similar phosphate-binding sites

Lineage:

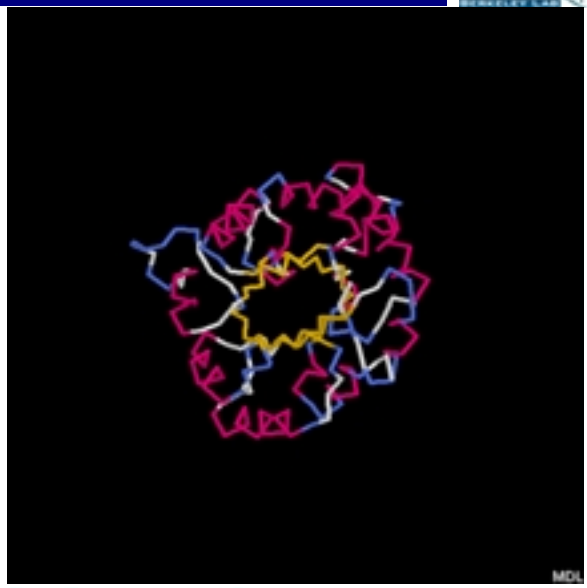
1. Root: scop
2. Class: Alpha and beta proteins (a/b)
   Mainly parallel beta sheets (beta-alpha-beta units)
3. Fold: TIM beta/alpha-barrel
   contains parallel beta-sheet barrel, closed; n=8, 2=8; strand order 12345678
   the first six superfamilies have similar phosphate-binding sites

Superfamilies:

1. Triosephosphate isomerase (TIM) (1)
2. Ribulose-phosphate binding barrel (3)
3. Thiamin phosphate synthase (1)
4. PMY-linked oxidoreductases (1)
5. Inosine monophosphate dehydrogenase (IMPDH) (1)
   The phosphate moiety of substrate binds in the "common" phosphate-binding site
6. PLP-binding barrel (2)
   circular permutation of the canonical fold: begins with an alpha helix and ends with a beta-strand
7. NAD(P)-linked oxidoreductase (1)
8. Glycosyltransferases (7)
9. Metallo-dependent hydrolases (3)
   the beta-sheet barrel is similarly distorted and capped by a C-terminal helix
   has transition metal ions bound inside the barrel
10. Aldolase (9)
    Common fold covers whole protein structure

SCOP Example: The TIM Barrel

The TIM Barrel is a common
Protein fold first observed
In triosephosphate isomerase



SCOP STATISTICS

Class	Number of folds	Number of families	Number of superfamilies
All Alpha Proteins	128	197	296
All Beta Proteins	87	158	251
Alpha and Beta Proteins (a/b)	93	153	323
Alpha and Beta Proteins (a+b)	168	237	345
Multi-domain Proteins	25	25	32
Membrane & Cell Surface Proteins	11	17	19
Small Proteins	52	72	102
Total	564	859	1368

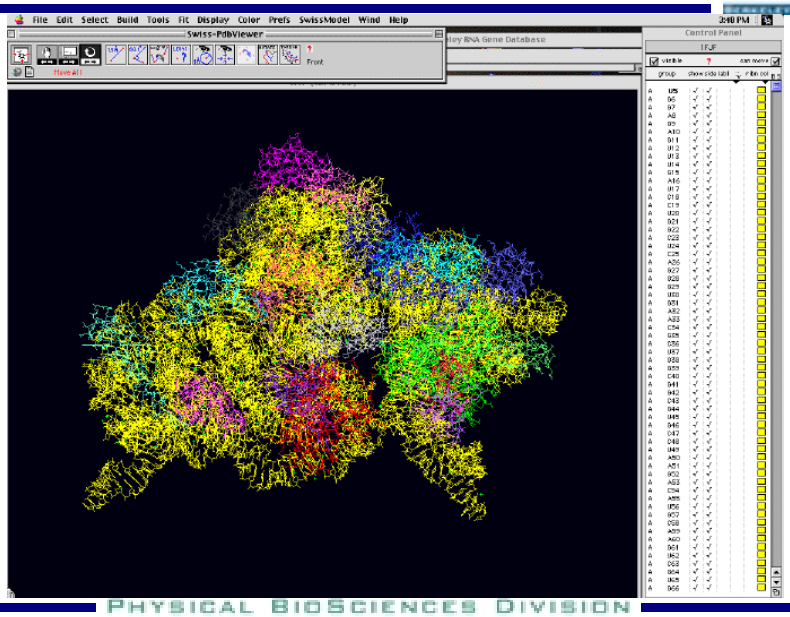
TOOLS AND DATABASES

http://www.expasy.ch/sprot/sprot-top.html

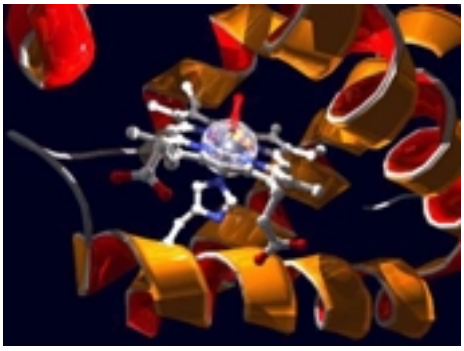
A Map to Tools and Databases

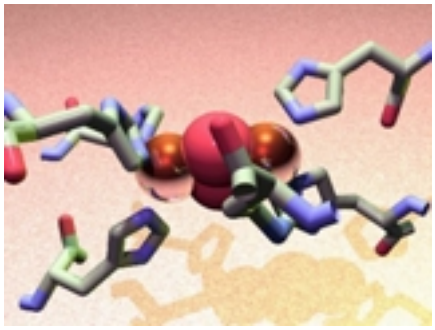
http://www.expasy.ch/sitemap.html

Swiss PDBViewer for Your PC (MAC)



SWISS PDBVIEWER GALLERY





BUILDING A PROTEIN MODEL

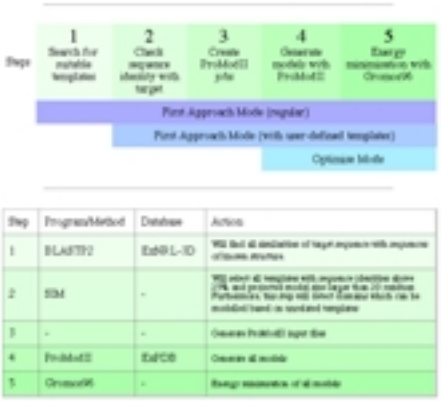
Building a Protein Model

How do you construct a model of a new protein from its amino acid sequence?

Procedure

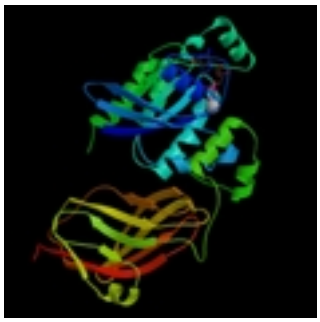
1. **Identify the closest related protein of known structure**
2. **From the alignment - replace the amino acid sidechains**
3. **For deletions - remove the deleted structure**
4. **For insertions - Add loops or other insertions using structures of similar sequence found in other proteins**
5. **Energy refine the modeled structure**

How SWISS-MODEL works

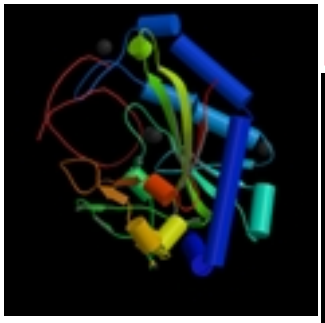


Protein Interactions

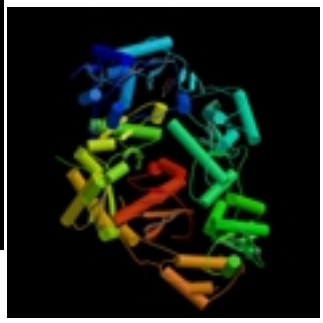
Protein-Protein Complex 1DS6
Crystal Structure Of A Rac-Rhogdi



Protein-DNA Complex 1DMU
Restriction Endonuclease Bgli Bound to its DNA Recognition Sequence



Protein-ligand Complex 1EET
HIV-1 Reverse Transcriptase In Complex With The Inhibitor Msc204



PROTEIN "DOCKING"

Docking: Computational Modeling of Protein complexes using known 3-D structures

Dock: Common Program

Step 1: Start with crystal coordinates of target receptor

Step 2: Generate molecular surface for receptor

Step 3: Generate spheres to fill the active site

Step 4: Matching

Sphere centers are matched to the ligand atoms, to determine possible orientations for the ligand. Typically tens of thousands of orientations generated for each ligand molecule.

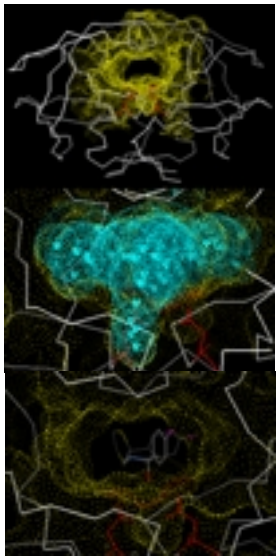
Step 5: Scoring

Each oriented molecule is scored for fit. Currently 3 scoring schemes:

Shape scoring

Electrostatic scoring

Force-field scoring



HIV-1 protease is the target receptor

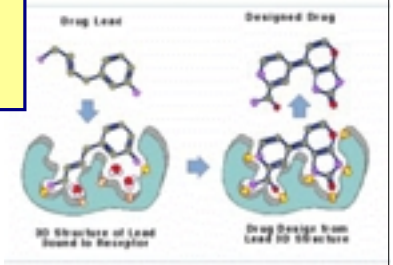
Blow up of the active site.

This is the top-scoring orientation for the molecule thioketal in the HIV-1-protease active site, using force-field scoring.

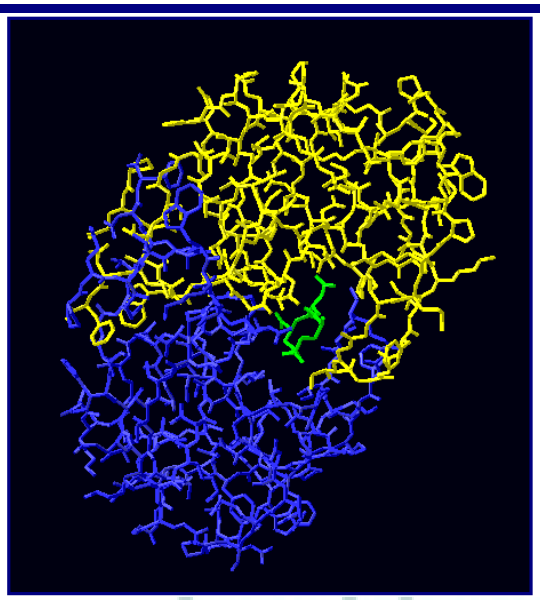
STRUCTURE-BASED DRUG DESIGN

Procedure for Structure-based Drug Design

- **Determine High-resolution structure of Protein target in complex with drug lead (ligand)**
- **From analysis of structure determine potential sites for modification to improve binding and other desired characteristics**
- **Synthesize a combinatorial library based on structure of complex (directed library)**
- **Screen for improved drugs**
- **Iterate (Determine structure of the protein complex with the designed drug)**



Structure-Based Drug Design



PDB file 1A30 displayed with Swiss PDB Viewer

HIV-1 protease with tripeptide inhibitor.

AIDS protease inhibitors are examples of successful applications of Structure-based drug design

Summary

- Structural Biology is a broad discipline that attempts to describe biological function in terms of 3-D MOLECULAR STRUCTURE
- High-throughput structure determination methods are being developed for crystallography and NMR - Structural genomics
- It is now possible to engineer proteins and nucleic acids for desired or improved functions
- Structure-based drug design is now yielding medically important pharmaceuticals